Spatial Sampling Design and Strategies

Jun Zhu

Department of Statistics and Department of Soil Science
University of Wisconsin - Madison
Madison, Wisconsin

Outline

Outline

Spatial sampling design

- Example: a study of old-growth northern hardwood forests (Miller et al., 2002).
  - Consideration of biodiversity in natural resource management.
  - Spatial patterns of forest understory vegetation (herbs, shrubs, tree seedlings, saplings).
  - Different species exhibit different spatial patterns within a given environment?
  - Biotic and abiotic factors in the environment are related to a species' spatial pattern?
- An important question: *where* should data be collected?
- The purpose is to design a sampling scheme that ensures scientific objectivity.

Spatial sampling design

- Suppose the study area of interest is D.
- Suppose measurements of Z will be taken at locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ in D, where $\mathbf{s} = (x, y)$ and n is the sample size. Where should they be?
- It depends!
- Possible objectives
  - Estimation of mean (e.g. average soil P in a field)
  - Estimation of variogram (e.g. map of soil P in the field)
  - Comparison of treatments (e.g. effect of a new fertilizer)
- Possible prior information
  - Accessible study area and sampling locations
  - Affordable sample size
  - Condition of a study area

Related subjects

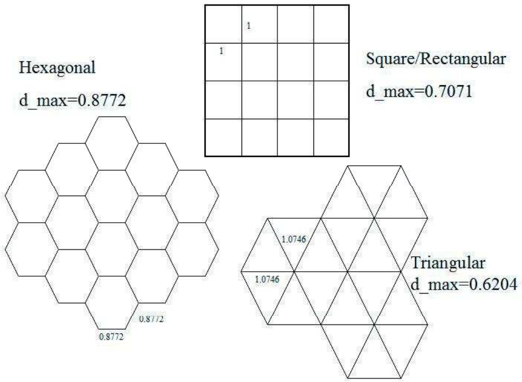
- Survey sampling: design-based sampling versus model-based sampling (Gruijter and Braak, 1990; Särndal et al., 1992)
- Design of experiment and optimal design (Mead et al., 1993)
- Spatial sampling design and optimal sampling (Webster and Oliver, 2001)
- An excellent review article: Stein and Christien (2003)

Outline

Regular grids

- Triangular or isometric grid: tiling plane regularly with equilateral triangles.
- Rectangular grid: tiling plane regularly with squares.
- Hexagonal grid: tiling plane regularly with hexagons.

Regular grids



$d_{\max}$ = maximum distance from any point in D to the nearest grid point.

A plausible scenario

- The goal is to estimate the overall mean

$$\mu = E(Z).$$

- Assume a regular spatial sampling grid with a fixed sampling density.
- Assume an exponential semivariogram for the spatial correlation structure.

And the winner is...

- A triangular grid is the most efficient design with the smallest $d_{\max}$.
- That is, for the same sampling intensity, it places the sampling locations as far apart as possible while minimizing the area that is under-represented.
- A triangular grid is most efficient for most bounded variograms that have finite ranges.

Remarks

- Under some other assumptions, a hexagonal grid is the most efficient design.
- For convenience, a rectangular grid is often the preferred design in field work.
- Major drawbacks of a regular grid include poor variogram estimates at short distances and the potential problems of systematic design (as versus randomized design).

Outline

Main idea

- To compensate for poor variogram estimates using regular grid designs, an improved method was proposed by Clayton and Hudelson (1995).
- The main idea is to use a regular grid system, but sample at unequal spacings.
- In one dimension (1D), the design allows the estimation of variogram at all multiples of the smallest lag with a minimum number of sampling locations.

1D transect

- Let $\times$ = sample; $\circ$ = skip (sampling).
- A 3/7 cyclic sampling design with 2 repeats looks like:

$\times$	$\times$	$\circ$	$\times$	$\circ$	$\circ$	$\circ$	$\times$	$\times$	$\circ$	$\times$	$\circ$	$\circ$	$\circ$

with lag distances

$\times$	1	-	3	-	-	-	7	-	-	-	-	-	-
-	$\times$	-	2	-	-	-	6	7	-	-	-	-	-
-	-	-	$\times$	-	-	-	4	5	-	7	-	-	-

1D transect

- Choice of specific sampling pattern is important.
- Why not

$\times$	$\times$	$\times$	$\circ$	$\circ$	$\circ$	$\circ$	$\times$	$\times$	$\times$	$\circ$	$\circ$	$\circ$	$\circ$

with lag distances

$\times$	1	2	-	-	-	-	7	-	-	-	-	-	-
-	$\times$	1	-	-	-	-	6	7	-	-	-	-	-
-	-	$\times$	-	-	-	-	5	6	7	-	-	-	-

- Lag distances 3 and 4 are missed.

Remarks

- For each lag distance, the proposed 3/7 design gives enough data for making the confidence intervals of the variogram small.
- There are more 7-lag distances than others in a 3/7 design, which cannot be avoided.
- Other possible cyclic sampling designs are: 4/13, 5/21, 6/31, etc. (Clinger and Ness, 1976).

2D region

- Extension to a 2D region is straightforward, but only approximately optimal.
- A 3/7 design for both the x-axis and y-axis:

× × ○ × ○ ○ ○ × × ○ × ○ ○ ○
× × ○ × ○ ○ ○ × × ○ × ○ ○ ○
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
× × ○ × ○ ○ ○ × × ○ × ○ ○ ○
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

- One can have different cyclic sampling designs for rows and columns.
- See Miller et al. (2002) for more details of the understory vegetation example.

Sampling design in practice

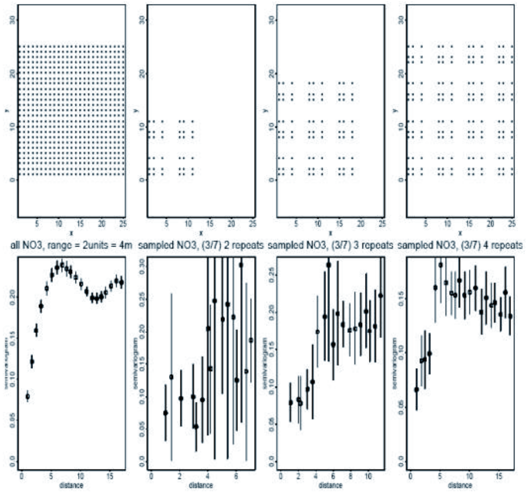
In practice, how to choose a particular cyclic sampling design and hence the sample size?

1. Conduct a pilot study to obtain a rough estimate of the range, sill, and nugget.
2. Simulate data on a grid with the finest grain scale possible for sampling, based on the estimated range, sill, and nugget.
3. Sample from the simulated data according to different sampling designs.
4. For each sample, compute the fitted range, sill, and nugget, and the confidence intervals of the variogram.
5. Evaluate the effect of different designs on the confidence interval width.
6. Consult a statistician!

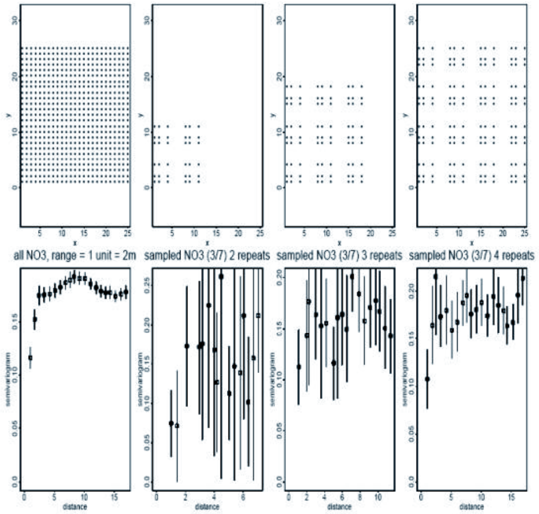
Example: Nitrogen cycling

- Assume exponential variogram model with
  - $r = 2$ (i.e. 95% effective $r = 6$).
  - $r = 1$ (i.e. 95% effective $r = 3$).
- Assume a 25×25 grid structure at a 2-m increment.
- Compare the use of 2D 3/7 cyclic sampling design with 1, 2, or 3 repeats.

$r = 2$



$r = 1$



Outline

Main idea

- In many field experiments, blocking is used to account for experimental unit (EU) heterogeneity, assuming that EUs within block homogeneous.
- Often there is spatial correlation within a block.
- If the goal is to have equal precision for the tests of treatment differences, it would make sense to design the experiment so that all treatments are equally near each other.

Example

Block	Treatment			
1	C	A	B	D
2	B	D	C	A
3	A	B	D	C
4	D	B	C	A

	Distance between plots				
Contrast	Block 1	Block 2	Block 3	Block 4	Average
A vs B	1	3	1	2	1.75
A vs C	1	1	3	1	1.50
A vs D	2	2	2	3	2.25
B vs C	2	2	2	1	1.75
B vs D	1	1	1	1	1.00
C vs D	3	1	1	2	1.75
Average					1.67

Average distance balanced design

- Not a balanced design since some treatments are on average closer than others.
- Simple switch in block 4 to DACB would result in much closer average distances.
- A strategy may be to strive for an average distance balanced design.

Nearest neighbor approach

- Instead of distance, look at neighbors of each treatment:
  A vs B: 2 A vs C: 3 A vs D: 0
  B vs C: 1 B vs D: 4 C vs D: 2
- Similar problem as before. While switching block 4 would help, we can do better.
- There are 12 neighbor pairs and 6 trt pairs:

Block	Treatment			
1	C	A	B	D
2	B	D	A	C
3	A	D	C	B
4	D	C	B	A

- Arrangement above is balanced for nearest neighbors and distance.
- Often correlation in both directions (2D). Similar approaches apply.

References

M. K. Clayton and B. D. Hudelson. Confidence intervals for autocorrelations based on cyclic samples. *Journal of the American Statistical Association*, 90:753–757, 1995.

W. Clinger and J. W. Van Ness. On unequally spaced time points in time series. *Annals of Statistics*, 4:736–745, 1976.

J. J. De Gruijter and C. J. F. Ter Braak. Model free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology*, 22:407–415, 1990.

R. Mead, R. N. Curnow, and A. M. Hasted. *Statistical Methods in Agriculture and Experimental Biology, 2nd Edition*. Chapman & Hall, London, 1993.

T. F. Miller, D. J. Mladenoff, and M. K. Clayton. Old-growth northern hardwood forests: Spatial autocorrelation and patterns of understory vegetation. *Ecological Monographs*, 72:487–503, 2002.

C. E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer, New York, 1992.

A. Stein and E. Christien. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment*, 94:31–47, 2003.

R. Webster and M. A. Oliver. *Geostatistics for Environmental Scientists*. Wiley, West Sussex, 2001.

I would like to thank. . .

- Bruce Craig, Purdue University
- Larry Douglass, University of Maryland - College Park
- Murray Clayton, University of Wisconsin - Madison
- Monica Turner, University of Wisconsin - Madison